Azure End to End Data Pipeline Project

Requirements

1. Functional Requirements
- Display key performance indicators (KPIs) including:
- Total products sold by gender and product category
- Total sales revenue by gender and product category
- Percentage split of male vs. female customers
- Filter dashboard views by:
- Product category
- Customer gender
- Date range (start date to end date)
- Support drill-down functionality for more granular insights.

2. Technical Requirements
- Build an automated ETL pipeline that:
  - Extracts data from the on-prem SQL database
  - Loads it into Azure Data Lake or Azure SQL Database
  - Applies necessary transformations (e.g., join tables, format dates, calculate aggregates)
  - Runs on a daily schedule
- Use Azure Data Factory or Azure Synapse Pipelines for orchestration.
- Store processed data in a format accessible to Power BI or similar visualization tools.
- Ensure data refresh and transformation logs are maintained for traceability.

3. Reporting Requirements
- Create a Power BI dashboard (or equivalent) that includes:
  - Visualizations for total sales, revenue, and gender split
  - Slicers/filters for product category, gender, and date range
  - Clean and intuitive user interface
  - Clear legends and tooltips to improve usability

4. Non-Functional Requirements
- Performance: Dashboards should load within 5 seconds for filtered queries.
- Security: Implement role-based access control for data views.
- Scalability: Support growing data volume without performance degradation.
- Maintainability: Modular ETL design to allow for easy future updates or extensions.
- Reliability: Daily pipeline runs should have alerting for failure or anomaly detection.